Published in final edited form as:

Drug Alcohol Depend. 2011 June 1; 115(3): 196–204. doi:10.1016/j.drugalcdep.2010.10.023.

Evaluating the validity and utility of scaling alcohol consumption indices alongside AUD symptoms in treatment-seeking adolescents

Bettina B. Hoepfner, Christopher W. Kahler, and Kristina M. Jackson

Center for Alcohol and Addiction Studies, Brown University, 60 Staniford Street, Boston, MA 02114 USA

Abstract

Background—Current initiatives to update diagnostic criteria for alcohol use disorders (AUDs) have stimulated dialogue about the usefulness of indicators of alcohol consumption in the diagnosis of AUDs.

Methods—This study used Rasch model analyses to examine the properties of alcohol consumption descriptors and AUD symptoms among 3,382 treatment-seeking adolescents, aged 12–18 years, in the DATOS-A (USDHHS, 1993–1995) baseline assessment, and evaluated the predictive validity of different scoring methods (with and without alcohol consumption) for 12-month alcohol involvement.

Results—Rasch model analyses supported the unidimensionality of indices of alcohol consumption and AUD symptoms. Test information functions showed that adding consumption items provides further information at all points of the alcohol involvement severity spectrum. Combining AUD symptoms with indices of alcohol consumption provided better prediction of alcohol involvement after treatment than either AUD symptoms counts or DSM-IV dependence diagnosis alone. Differential item functioning (DIF), however, was observed for select items. Generally, indices of drinking “too much too fast” were more severe for females, African Americans and Hispanics, while the opposite was true for items measuring “too much too often”. For age, “too much too often” items were more severe for the younger (12–14yrs) age group, and AUD symptoms were more severe for the older (15–18yrs) age group.

Conclusions—Indices of alcohol consumption can be validly scaled along with AUD symptoms in this population, and their inclusion provides statistical measurement advantages. Nevertheless, caution is necessary in using consumption items in measuring alcohol involvement due to DIF observed across sex, race and age.

Keywords

Alcohol Consumption; DSM-IV; Rasch Modeling; Adolescence; Measurement

Corresponding Author: Bettina B. Hoepfner, Ph.D., Center for Addiction Medicine, Massachusetts General Hospital, 60 Staniford Street, Boston, MA 02114, Phone: 401-863-6630, Fax: 401-863-6647, bhoepfner@partners.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Alcohol use disorders (AUDs) are defined as maladaptive patterns of alcohol use, which lead to clinically significant impairment or distress. Absent from diagnostic criteria, however, are specific patterns of drinking or indicators of alcohol consumption. Saha and colleagues (2007) noted that historically, the abandonment of alcohol consumption indices as diagnostic criteria can be backdated to the 1960s, when self-reported quantity/frequency of drinking in a study on criminals could not be validated against the reports of their relatives, who did not know this information (Guze et al., 1963). Research diagnostic criteria appearing after 1971 no longer included alcohol consumption indicators for any AUDs (Saha et al., 2007), including current editions of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV, American Psychiatric Association, 1994) and the *International Classification of Diseases* (ICD-10, World Health Organization, 1992).

In the past 3 decades, significant advances have been made in the measurement of alcohol consumption (Dawson, 2003; Greenfield and Kerr, 2008), and both the APA and the WHO have initiatives underway to update diagnostic criteria of AUDs. These efforts have stimulated dialogue about the usefulness of indicators of alcohol consumption in diagnosis, treatment, and prevention of AUDs (Li et al., 2007). Past research based on large epidemiological datasets has indicated that DSM-IV alcohol abuse and dependence criteria and associated symptom items can be scaled validly along a single underlying continuum of alcohol involvement, but that there is a scarcity of criteria tapping the less severe range of the continuum (Kahler and Strong, 2006; Proudfoot et al., 2006; Saha et al., 2006). Results from general population item response theory (IRT) studies have indicated that consumption indicators may be well suited to measure this region (Kahler et al., 2008; Saha et al., 2006; Saha et al., 2007).

This paper seeks to evaluate the scaling of indices of alcohol consumption alongside diagnostic criteria for AUDs in a sample of substance-using, treatment-seeking adolescents. Similar to research findings in adults, research on adolescents has shown that abuse and dependence criteria can be scaled along a unifying continuum (Chung and Martin, 2001; Gelhorn et al., 2008; Harford et al., 2009; Martin et al., 2006). Unlike adults, however, adolescents and young adults undergo normative developmental changes, which may affect clinical significance and symptom prevalence (Achenbach, 2009). For example, tolerance (Kahler and Strong, 2006) and craving (Bucholz, 2009) indicators have different severity interpretations in different age groups (e.g., young adults vs. adults). Additionally, specific indicators may be perceived by different age groups as having different meaning. For example, Chung et al. (2005) noted that the criterion “drinking more or longer than intended” includes the embedded assumption of individuals setting a limit of use, but that adolescents do not typically set such a limit.

In both adults and adolescents, the relationship of indices of alcohol consumption to AUD symptoms has been rarely studied. Current research efforts have focused on binge drinking, where the scaling of different frequencies of consuming 5+/4+ (for men and women, respectively) drinks has been examined (Saha et al., 2007). Results indicate that drinking 5+/4+ drinks at least once a week may be a suitable criterion for future classifications of AUDs (Saha et al., 2007), but that the inclusion of weekly 5+/4+ drinking in diagnostic reformulations would also result in lower rates of diagnosis for Blacks, Hispanics and women (Keyes et al., 2009). Other indices of alcohol consumption (e.g., drinking on a weekly basis, drinking more than 14(male)/7(female) drinks per week) have also been shown to fit on the same dimension as AUD symptoms and to map less severe regions of the continuum (Kahler et al., 2008). Most recently, both the frequency of at-risk drinking and the scaling of the quantity of at-risk drinking have been examined in a cross-national study

of emergency department patients (Borges et al., 2010). Here, drinking 5+ and 12+ drinks were examined, which were found to be useful in tapping the lower and middle part of the alcohol involvement severity, as Saha et al. (2007) originally suggested. It should be noted that each of these studies has focused on adults.

Only one study that we are aware of has addressed the scaling of alcohol consumption in adolescents. An examination of narrowly defined age groups cross-sectionally and longitudinally (i.e., 12–14yrs and 15–18yrs, followed up 6 years later) using data from the National Longitudinal Study of Adolescent Health (Add Health, Udry, 2003), showed that alcohol consumption can be scaled validly on a continuum of alcohol involvement with alcohol problems (Kahler et al., 2009). Indices of alcohol involvement functioned similarly across sex and race ethnicity. Age-related differences, however, were found. Items such as “gotten drunk” and “drinking on a monthly basis” were more severe for younger respondents whereas alcohol problems such as “school problems due to alcohol” were more severe for older youth. These findings suggest that the scaling of alcohol consumption may be sensitive to developmental changes.

The sample used in the Kahler et al. (2009) study addressed alcohol involvement in the general population of adolescents, and as such, was unable to evaluate the scaling of more severe alcohol consumption indices or their relationship to AUD symptoms. The present study draws on a large national sample of substance-using, treatment-seeking youth. Use of a clinical sample permits examination of distinct low-base-rate, high-severity indices of alcohol involvement, including consumption that reflects very heavy and risky alcohol consumption. We used an IRT approach to evaluate alcohol consumption indices that characterize both drinking “too much too fast” (i.e., various definitions of heavy drinking episodes) and “too much too often” (e.g., various scaling of regularity of 5+ drinks), as recommended by Li et al. (2007). Based on previous research on adults (Borges et al., 2010) and adolescents (Kahler et al., 2009), we expected lifetime indices of drinking “too much too fast” to provide information on the lower end of the alcohol severity continuum (consistent with the relatively high prevalence of heavy episodic drinking; Johnston et al., 2009), while we expected indices of “too much too often” to be indices of greater severity, particularly as frequency increases. We assessed whether the inclusion of alcohol consumption indices with AUD symptoms is valid and quantified the extent to which their inclusion represents a statistical measurement advantage in terms of measurement precision and predictive validity. We hypothesized that combining AUD symptoms with indices of alcohol consumption would provide more information about alcohol involvement severity and better prediction of future alcohol involvement than either AUD symptoms counts or DSM-IV dependence diagnosis alone.

2. Methods

Data were from the Drug Abuse Treatment Outcome Study for Adolescents, DATOS-A (United States Department of Health and Human Services, 1993–1995), a prospective, community-based study of adolescents entering treatment designed to evaluate the effectiveness of adolescent drug treatment. In-person interviews were conducted at intake, at 1, 3, and 6 months during treatment, and (for 1,785 respondents) 12 months after treatment termination. Data for the present study are taken from the intake and 12-month follow-up assessments.

2.1 Participants

Participants were $n=3,382$ treatment-seeking adolescents between 12–18 years of age, who were sampled from among those admitted to treatment in 1993–1995. Participants included treatment seekers for a variety of substance use problems, which included alcohol for some

but not all respondents¹. At baseline, 25.4% of the sample received a DSM-III diagnosis of alcohol dependence (including mild, moderate and severe dependence), and 10.1% received a DSM-III “abuse only” designation. Seventeen percent reported no prior alcohol use.

The sample was predominantly male (73.8%), 51.6% White, 23.9% African American, and 20.5% Hispanic, with 4.0% using other categories to describe their racial background. The average age was 15.8 (standard deviation (SD) = 1.4) with a median age of 16. For the purposes of this paper, two age groups were differentiated: early adolescents (aged 12–14 years, mean=13.7, SD=0.6), who comprised 18.6% ($n=629$) of the sample, and older adolescents (aged 15–18 years, mean=16.2, SD=1.0), who comprised 81.4% ($n=2753$) of the sample. Age 14 was used as the cut point, because it corresponds to the age at which youth transition from middle to high school. This full sample was used for IRT analyses.

At the 12-month follow-up interview, 2,974 of the clients who completed the intake interview were selected as eligible for the 12-month follow-up interview (reasons for ineligibility include incarceration, hospitalization, and living outside of the metropolitan interviewing area). Of these eligible clients, 65.9 % ($n = 1,960$) were located, 60% ($n = 1,785$) were successfully interviewed, 0.5% ($n = 14$) were deceased, and 5.4% ($n = 161$) refused to participate (USDHHS, 1993–1995). In other words, 91.1% of eligible and located clients were interviewed. This sample was used to test the predictive validity of different scoring methods.

2.2 Measures of lifetime alcohol involvement

This study focuses on items assessing lifetime indices of alcohol consumption and AUD symptoms in order to make the results of this study more readily comparable to findings of other IRT studies on clinical adolescent populations, all of which have analyzed lifetime use (Chung and Martin, 2001; Gelhorn et al., 2008; Martin et al., 2006). More generally, a lifetime framework is also desirable because diagnostic criteria need to make sense within a larger framework, which includes genetic and epidemiological inquiry (Hasin and Beseler, 2009).

2.2.1 Alcohol consumption—Seven indices of alcohol consumption were available in the DATOS-A dataset. Consumption items addressing “too much too fast” were “drunk”, and “most drinks in a row” (coded as 5+ and 12+). “Too much too often” was assessed using the items “5+ drinks per day once a week for 2+ months”, “drunk for several days in a row”, and “5+ drinks every day for 2+ weeks”. As an additional measure of at-risk weekly drinking, we also included an item measuring drinking 14+/7+ drinks in a week for males and females respectively, as defined by the National Institute on Alcohol Abuse and Alcoholism (NIAAA, 2009) as at-risk drinking.

2.2.2 AUD symptoms—A variety of indices of alcohol dependence and abuse were used. All seven DSM-IV dependence criteria were assessed, including “tolerance”, “withdrawal”, “larger/longer”, “quit/control”, “time spent”, “activities given up” and “physical/psychological problems”. Of these, “withdrawal” and “physical/psychological problems” were assessed via multiple questions rather than through a single item in DATOS-A, and thus we used testlets to score them. Namely, “physical/psychological problems” was assessed using four items in DATOS-A: “Did you continue to use [alcohol] after you realized it was causing you any of these emotional problems?”, “Did you continue to drink after you realized that it was causing you any of these health problems?”, “Did you drink alcohol on more than one occasion after you realized it was causing you these health

¹The publically available DATOS-A dataset does not include information about participants’ primary reason for seeking treatment.

problems?”, and “Have you ever continued to drink when you knew you had any (other) serious physical illness that might be made worse by drinking?”. “Withdrawal” was assessed using three items: “Did stopping or cutting down on drinking alcohol ever cause you problems such as the shakes (hands tremble), being unable to sleep, feeling anxious or depressed, sweating, your heart beating faster, stomach aches, headaches, weakness, seeing or hearing things that weren’t really there, fits or seizures after stopping or cutting down on drinking, or the d.t.’s (delirium tremens)?”, “Have you had problems like these several times when you stopped or cut down on your drinking?”, and “Did you ever take a drink to keep from having problems like these or to make them go away?”. A positive response to any one item within either testlet was coded as an endorsement of the symptom, resulting in one binary indicator for each of the two testlets.

Three additional dependence items were included in DATOS-A which address other dimensions of dependence. Two of these items tap into Edwards and Gross (1976)’s criteria for the alcohol dependence syndrome. Namely, “difficulty stopping before becoming intoxicated” addresses loss of control, and “needed so badly that you could not think of anything else” addresses both the increased salience and craving aspect of dependence. The third item, “made sure you had alcohol available over 1+ month(s)”, is not part of the DSM nor part of Edwards and Gross’s description, but this type of indicator has been shown to be a high performing item in assessing alcohol dependence in clinical adult populations (Kahler et al., 2003).

Only two (DSM-IV) abuse diagnostic criteria were included: “social/interpersonal problems” and “hazardous use”. The DATOS-A questionnaire, which was based on DSM-III (APA 1981) rather than DSM-IV (APA 1994), combined “failure to fulfill role obligations” and “recurrent alcohol related legal problems” in one item, which furthermore did not fully address recurrence, and thus these two diagnostic criteria could not be included.

2.3 Measures of alcohol involvement at 12-month follow-up

To assess predictive validity, we examined three indicators of varying levels of alcohol involvement severity at 12-month follow-up: “drunk any alcohol during the past 12 months” (endorsed by 68.1% of the retained sample), “number of days drunk in the past 12-months” (mean = 24.7, SD = 52.0), and “drunk for several days in a row in the past 12 months” (endorsed by 6.0%).

2.4 Analytic Strategy

We used item response modeling based on the one-parameter logistic (1-PL) Rasch model (Birnbaum, 1968; Lord, 1980; Rasch, 1960; Wright and Masters, 1982) as the guiding psychometric framework. This model assumes that the discrimination parameter (i.e., the ability of an item to discriminate between individuals above vs. below the item’s severity threshold) is equal across items, which has practical advantages in the interpretation of severity ratings (i.e., they have the same meaning relative to each other, unlike severity estimates obtained in 2-PL models (Salzberger, 2002)), which are of primary interest here. The idea of equal item discrimination is also consistent with the common practice of sum-scoring items. Although by definition 2-PL models fit better than 1-PL models, severity estimates correlate very highly between 1-PL and 2-PL severity estimates (e.g., $r=0.99$ in Kahler et al., 2009; consistent with Wright, 1995)². Moreover, information contained in the discrimination parameter is largely contained in item infit statistics, which we report. Thus, we selected the 1-PL model to maximize interpretability.

²In this study, the correlation of the 1-PL severity estimates reported here and the severity estimates obtained through the use of a 2-PL model was 0.97 in the full sample analysis.

2.4.1 Unidimensionality—We tested the unidimensionality of the items (AUD symptoms and consumption indices) to determine whether they can be modeled validly as indicators of a single unidimensional latent trait. To this end, we used exploratory principal component analysis (PCA) based on the tetrachoric correlations of the items, as recommended by Cook et al. (2009). To determine the number of components to be retained, we used Horn's (1965) test of Parallel Analysis, which has consistently been shown to be among the most accurate procedures for determining the number of components to be retained in exploratory factor analysis (O'Connor, 2000; Zwick and Velicer, 1986). Rasch models were then used to scale the severity of both items and persons along an underlying latent continuum using an equal interval logit or log odds scale, and unidimensionality was once again evaluated using Rasch model indices. Specifically, we compared the estimated accounted variance to the expected variance if the data fit the Rasch model perfectly, and interpreted high discrepancy between empirical and expected variance as indicating lack of fit of unidimensionality (Linacre, 2006). Rasch model residuals were examined using a PCA, where a 2nd factor with an eigenvalue larger than 1.5 is indicative of potential secondary dimensions in the data (Linacre, 1998; Smith and Miao, 1994). Finally, the specific fit of individual items to the Rasch model were examined using two model fit indices: the infit mean square error (MSQ) of the items was judged to represent adequate fit, if values ranged from 0.60 to 1.40, which represents a norm for survey rating scales such as this one (Linacre and Wright, 1994). Point biserial correlations (PTBS; i.e., correlation between the responses to an item by each person and the total marginal score) were used to evaluate the correspondence of the individual items to the overall scale, where higher values reflect greater correspondence.

2.4.2 Differential item functioning (DIF)—Given that there are differences in alcohol consumption between boys and girls across race/ethnicity (Johnston et al., 2009) and prior IRT studies have demonstrated age differences in consumption indices, we tested for DIF due to demographics (sex, race, age) using Mantel-Haenszel *t*-tests (Mantel and Haenszel, 1959). Given the large sample and therefore high statistical power of these analyses, we utilized minimum effect sizes as a criterion for identifying DIF, so as to identify not just statistically significant but also substantively meaningful DIF: DIF > 0.50 denoted meaningful differences (Draba, 1977) and DIF > 0.64 denoted “moderate to large effects”, a standard used by the Education Testing Services (Tristán, 2006).

2.4.3 Test information—With fit to the Rasch model, an unweighted sum of items (i.e., 0 for lack of endorsement, 1 for endorsement of the item) can serve as a sufficient measure of an individual's severity, which is consistent with how items are typically used in scales; unweighted sums have also been shown to perform as well as IRT-derived weighted scale scores (Dawson et al., 2010). To compare the information contained in alcohol involvement scores with and without consumption indicators, we created two sum scores: the sum of all AUD symptoms, and the sum of AUD symptoms and consumption indices. For both sets of items (i.e., “tests”), we graphed test information functions, which are mathematical expressions of how much information a set of items contains (i.e., the height of the curve) over a range of severity values (i.e., the width of the curve). Test information curves are influenced by the number of items contained in the “test”, where more items generally though not always lead to higher test information. Unlike reliability estimates, however, test information functions show for which part of the continuum the “test” is most informative.

2.4.4 Prediction of future alcohol involvement—Using both sets of items, as well as a binary variable denoting DSM-IV dependence that was computed from the seven dependence items, we predicted 12-month alcohol involvement indices of varying severity. We used logistic regression for binary outcomes and Poisson regression for count data. Because only a subset of participants was successfully interviewed 12-months later, we used

logistic regression to determine differential response biases due to sex, race and age. If statistically significant, we included them as covariates in the predictive validity analyses to control for response bias. To rule out that 12-month interview completion was confounded with the baseline alcohol involvement scores we were interested in comparing, we also tested if they predicted 12-month survey completion. Finally, we included a covariate capturing whether participants spent any time in controlled environments during the past 12 months, which would limit their access to alcohol. This variable was scored if respondents endorsed time spent in jail, prison, or a juvenile detention home or time spent in short-term in-patient or long-term residential treatment units, or admittance to a halfway house for alcohol or drug problems. To enable comparison across models, alcohol involvement predictor variables were standardized (i.e., mean=0, SD=1) prior to analysis.

3. Results

3.1 Validity of Unidimensionality

In addition to items that were subsets of each other (i.e., “most drinks”, which was coded for 5+ and 12+), and thus did not have valid tetrachoric correlations, two highly correlated items (i.e., “5+ drinks per day once a week for 2+ months “ was correlated with “5+ drinks every day for 2 weeks” at $r=0.95$; “drunk” was correlated with “most drinks in a row: 12+” at $r=0.87$) led to a non-positive definite correlation matrix. To enable PCA, one variable of each pair was retained. The remaining 16 items yielded a clear 1-factor structure, with the first factor accounting for 67.8% of the variance (eigenvalue=10.6), and the second factor accounting for only 6.1% (eigenvalue of 0.98). By comparison, the average expected eigenvalue for the second factor based on random chance using Parallel Analysis (Horn, 1965) was estimated to be 1.10.

For the Rasch models (i.e., total sample and demographic subgroup analyses), all 19 items were included. Explained variance estimates were extremely similar to expected estimates (i.e., 56.7% empirical vs. 56.1% expected), indicating good fit of the Rasch model. Likewise, PCA of the model residuals estimated the eigenvalue of a 2nd factor at 1.4, also indicating unidimensionality.

Table 1 shows the Rasch model parameters. Individual item indices identified only 1 item with less than desirable model fit: the item “made sure you had alcohol available over 1+ months” was outside the recommended MSQ range with an MSQ value of 1.46. All other items were well within the specified range of 0.60 to 1.40. Pointbiserial correlations ranged from 0.32 (i.e., the same item flagged by the MSQ value) to 0.89, and were satisfactory.

3.2 Differential Item Functioning

In Table 1, items with DIF are flagged, and their item severity per group is presented in Table 2. Of the 20 instances of significant DIF between subgroups, 13 involved consumption items with 4 items accounting for 12 instances of DIF. Alcohol consumption DIF due to sex and race shared a pattern, where high volume drinking on one occasion (5+ and/or 12+) was more severe for females than males, African Americans and Hispanics than Whites, and Blacks than Hispanics, while the opposite was true for “14+/7+ drinks per week” and “5+ drinks per day once a week for 2+ months”. Among the AUD symptoms, “needed so badly that you could not think of anything else” was more severe for males. For race DIF, the dependence criterion “larger/longer” was more severe for Hispanics than Whites and the DSM-IV abuse criterion “hazardous use” was more severe for African Americans than Whites.

Age DIF followed a different pattern. For all items with age DIF, consumption items (i.e., “5+ drinks per day once a week for 2+ months”, and “5+ drinks every day for 2 weeks”)

were more severe for the younger age group, aged 12–14, while dependence items (i.e., “larger/longer”, “quit/control”, and “withdrawal”) were more severe for the older age group, aged 15–18.

The raw sum scores and the Rasch-model derived severity estimates corresponding with each score are plotted in Figure 1. For each subgroup, severity estimates of each score (i.e., 0–19) are plotted and compared within sex (A), race (B) and age (C). For all three comparisons, discrepancies in the severity of total test scores are most noticeable for low alcohol involvement scores. Alignment is best around a score of 4, which corresponds roughly to a severity of –1.5, ranging from –1.49 for Whites to –1.70 for African Americans.

3.3 Test Information With and Without Alcohol Consumption Items

The test information functions for two sets of items are graphed in Figure 2: using AUD symptoms only, and using both AUD symptoms and consumption items. It is evident that the test information function is higher for the set of items that includes consumption items at all levels of alcohol severity. Alcohol consumption items provide further information at high levels of alcohol severity by measuring gaps between regions measured by AUD symptoms, and provide information on the lower end of alcohol severity, which is not measured by AUD symptoms.

3.4 Predictive Validity

The logistic regression analysis indicated that none of the three compared alcohol involvement scores (i.e., sum score of AUD criteria only, sum score of AUD and consumption indices, and binary variable denoting DSM-IV alcohol dependence) were related to whether baseline participants ($n=3,382$) were successfully interviewed 12 months later ($n=1,785$). In testing demographics biases, results showed that sex (Wald $\chi^2=17.1$, $p<0.001$) and race (Wald $\chi^2=54.7$, $p<0.001$) but not age statistically significantly predicted 12-month interview completion. Specifically, males (50.2% successfully interviewed vs. 60.0% of females), African Americans (46.3% successfully interviewed vs. 58.4% of Whites) and Hispanics (44.2% successfully interviewed) were less likely to be successfully interviewed. Even though age was not a statistically significant predictor, we decided to include age alongside sex and race as covariates in the predictive validity analyses, so that our use of all three demographic variables would be consistent within each type of analysis (i.e., IRT models and predictive validity analyses).

Table 3 provides the parameter estimates, test statistics and effect sizes of the three different scoring methods in predicting 12-month alcohol involvement. For all three dependent variables, using a sum score of all items (i.e., AUD and consumption items) resulted in the best model, as indicated by the highest Pseudo R^2 values. In fact, in predicting the 12-month alcohol involvement descriptor of the lowest severity (i.e., “any alcohol in the past 12 months”), only the sum score of AUD and alcohol consumption indices was a statistically significant predictor, while the sum score of AUD diagnostic criteria only and the dichotomous indicator of DSM-IV dependence were not. Additionally, using a sum score of AUD symptoms explained more variance than using a dichotomous indicator of DSM-IV dependence for all three 12-month alcohol involvement indices, though only trivially so for the most severe alcohol use outcome tested.

Discussion

This study examined AUD symptoms and alcohol consumption indices in a clinical sample of adolescents with varying degrees of alcohol involvement. It extends previous research in

this area through its focus on adolescents and its extensive consideration of alcohol consumption indicators. Previous research has focused largely on the 5+/4+ conceptualization of heavy episodic drinking (e.g., Hasin and Beseler, 2009; Saha et al., 2007). The present study evaluated additional indicators of consuming “too much too fast” (e.g., 12+ drinks), and also included a variety of indicators of consuming “too much too often” (e.g., 5+ drinks per day once a week for 2+ months, 5+ drinks per day every day for 2+ weeks). Our findings indicate that alcohol consumption and AUD symptoms can be scaled validly along the same continuum of alcohol involvement. This conclusion rests on the convergence of several methods used to test the unidimensionality of the items.

Our results further indicate that alcohol consumption indicators also contribute measurement information throughout the continuum of alcohol involvement, ranging from low to high severity. Consumption indicators are interspersed with AUD symptoms throughout the alcohol involvement continuum, with indices of consuming “too much too fast” largely populating the lower range of severity, and indices of consuming “too much too often” measuring both medium (i.e., “drunk for several days in a row”, severity estimate = 1.14) and high severity (i.e., 5+ drinks per day every day for 2+ weeks, severity estimate = 1.62), rivaling and exceeding the severity of abuse and dependence DSM-IV diagnostic criteria.

In examining the predictive validity of using indices of alcohol consumption, we found that a combined sum score of consumption and AUD items accounted for more variance than either a DSM-IV dependence indicator, or a sum score based solely on AUD symptoms. This finding was consistent across three indices of 12-month alcohol involvement of varying degrees of severity. Thus, we conclude that using alcohol consumption indices in addition to AUD symptoms represents a statistical measurement advantage in terms of measurement precision and predictive validity.

Our results also, however, point to the need for caution in using consumption items in measuring alcohol involvement. Most of the instances of differential item functioning observed in this study, across sex, race and age, were associated with consumption items. For sex and race, there appeared to be a split, where consumption items measuring “too much too fast” were more severe for females, African Americans and Hispanics, while the opposite was true for items measuring “too much too often” patterns of consumption. In a general population sample of adolescents, Kahler et al. (2009) found that the items “gotten drunk” and “5 drinks in a row” were more severe for African Americans and/or Hispanics, and as such, our findings about consuming alcohol “too much too fast” are in line with previous research. The Kahler et al. (2009) study did not, however, include indicators of consuming “too much too often”, and thus our findings cannot not be compared to that study. In adults, Saha et al. (2007) examined the scaling of drinking 5+/4+ drinks and found no sex or race differences. Thus, while our findings of differentiating effects between “too much too fast” and “too much too often” indicators for sex and race in this clinical adolescent population are suggestive of conceptual differences, they represent a first and isolated observation.

Importantly, the pattern of differential item function in alcohol consumption items was different for age than sex and race. For age, there was a noticeable split between “too much too often” indicators of consumption and AUD symptoms, where the “too much too often” items were more severe for the younger (12–14yrs) age group, and AUD symptoms were more severe for the older (15–18yrs) age group. It is further evident from the data that this split does not represent an artifact of a developmental shift where low severity indicators become more normative with increasing age, because both types of items included low and high severity indicators. As such, this finding suggests the presence of qualitative age differences in the severity of AUD symptoms and “too much too often” consumption

indicators even though both sets of indicators tap a common underlying continuum; intuitively, it makes sense that both higher alcohol consumption and more dependence criteria are associated with greater alcohol involvement but that consumption items are distinct from rather than redundant with dependence items. It is in contrast to findings from clinical adult populations for drinking 5+ drinks and AUD symptoms, which found no age differences between young adults (18–29) and older adults (30–44, 45+) (Saha et al., 2007). This difference in findings speaks to the importance of evaluating these indices in adolescents in addition to adults. Given the rapid pace with which drinking behavior changes in younger ages, evaluating age differences in a more fine-grained manner is especially important. Previous IRT studies on clinical adolescent samples (Gelhorn et al., 2008; Martin et al., 2006) did not examine consumption items describing either “too much too often” or “too much too fast” indicators, nor did they examine age differences within adolescence, and thus more research is necessary to evaluate this finding.

We also identified differential item functioning for DSM-IV criteria for sex and race. In their IRT analysis on a clinical sample of adolescents, Martin et al. (2006) noted that gender differences existed in thresholds for hazardous use, legal problems, and physical-psychological problems. Similarly, our own results noted differential item function for AUD symptoms, but differed in their identification of specific items. We found differential item functioning for hazardous use for race (i.e., more severe for African Americans), but not sex. Withdrawal symptoms appeared to function differently by sex (i.e., more severe for males), but physical or psychological problems did not. Unfortunately, Gelhorn et al. (2008) did not assess the effect of sex and race in their adolescent sample, and thus this discrepancy cannot be resolved through the consideration of this third set of data.

In sum, alcohol consumption indices appeared to be more prone to differential item functioning than AUD diagnostic criteria across sex, race and age. On a total test score level, these differences are minimal. In selecting specific items for measuring alcohol involvement, however, our results suggest that critical attention needs to be paid to the nature of the items to be included versus excluded, because indices of drinking “too much too fast” and “too much too often” function differentially in sex, race and age subgroups. Specifically, drinking “too much too often” appeared to be less severe for African Americans and Hispanics, which is in line with the finding that the inclusion of weekly 5+ drinks in diagnostic reformulations would result in lower rates of diagnosis for Blacks and Hispanics in adult populations (Keyes et al., 2009). Indeed, the amount of observed DIF cautions against including consumption indices in revised AUD diagnostic criteria, at least for adolescent samples. For measuring alcohol involvement, however, the addition of alcohol consumption items, when balanced between “too much too fast” and “too much too often”, can add measurement precision and predictive power.

Of note is also our finding regarding the item “made sure you had alcohol available over 1+ month(s)”. In adults, this non-DSM-IV dependence indicator was a very high performing item (Kahler et al., 2003), although the item in that study did not assess duration of ensuring alcohol’s availability. In this clinical adolescent sample, however, the item performed unexpectedly poorly, indicating that it may work differently for adolescents versus adults, perhaps due to limited alcohol availability for adolescents versus adults. Two other alternative indicators of dependence (one related to loss of control and the other to craving) performed well as indicators of alcohol involvement severity but tended to map regions of the continuum that were already well covered by DSM dependence criteria. Thus, it is not clear on statistical grounds that adding such criteria would enhance coverage of the dependence continuum.

4.1 Strengths and Limitations

The present study drew on a large clinical sample of treatment-seeking adolescents, and is the first to examine the scaling of alcohol consumption alongside AUD symptoms in this age group. Our focus on a clinical sample of adolescents enabled us to examine the fit of low base-rate, high severity measures, which couldn't be examined in a general population study of adolescents (Kahler et al., 2009). Only a handful of studies have considered a measure of alcohol consumption as a possible diagnostic criterion, and have focused exclusively on heavy episodic drinking. In this study, we examined seven consumption indicators, including heavy episodic drinking as well as measures of drinking too much too often, the latter of which type of consumption has not been previously examined in this context.

Some limitations must be noted. Importantly, all of the data presented here represent self-reports. In the past, the validity of self-reports of alcohol consumption in particular has been of concern. Two issues are at stake: the honesty with which people report versus the accuracy with which they are able to do so. Honesty may be impacted by social desirability, but would be similar for AUD diagnostic criteria and consumption indices. Accuracy may be impacted by recall biases. Comparisons of different types of alcohol reports, however, have consistently shown that while self-reports with shorter (e.g., 7-day) or no recall periods (e.g., ecological momentary assessment) tend to result in higher drinking estimates than those with longer (e.g., 30-day) recall periods, such differences are modest (Carney et al., 1998; Hoepfner et al., in press; Maisto et al., 2008; Perrine et al., 1995; Toll et al., 2006). In this study, we used a binary coding ("ever" vs. "never") of specific drinking events, which should be less impacted by recall biases, but recall periods were also long. Thus, the reported alcohol consumption is likely an underestimate of actual consumption.

Additionally, the data used in this study were collected 15 years ago. A recent study (Faden and Fay, 2004), however, showed that in this age group drinking prevalence rates of a variety of indicators (i.e., any use in the past 30 days, 5 drinks in a row in past 2 weeks, and daily use in past 30 days) have remained relatively stable throughout the 5 to 10 years previous to that study (i.e., 1994–2004), including the time during which the DATOS data were collected. This finding suggests that though the DATOS data are not current, they are likely similar to current trends. Relatedly, however, the DATOS-A interviews, upon which these data are based, were structured to assess DSM-III criteria. While there is considerable overlap with DSM-IV criteria, not all DSM-IV criteria were assessed. Also of relevance is the fact that the studied sample was not confined to adolescents seeking treatment for alcohol. Rather, it included adolescents who were seeking treatment for any type of substance use, including alcohol.

Finally, it should be kept in mind that the predictive validity analyses are limited in their generalizability due to the 53% rate of 12-month follow-up completion. Our attrition analyses showed that the completion of the 12-month follow-up interview was not predicted by the types of alcohol involvement scores we were interested in comparing (i.e., sum score of AUD criteria only, sum score of AUD and consumption indices, and binary variable denoting DSM-IV alcohol dependence) in terms of their predictive validity. That finding suggests that the predictive validity analyses were likely unbiased by differential drop out by participants with high alcohol involvement. The 53% rate of 12-month follow-up completion nevertheless limits generalizability. We identified demographic biases in 12-month interview completion, and accounted for them in the predictive validity analyses. There are other factors that may also be related to drop-out which we did not include in our model, and they limit the generalizability of the 12-month outcome findings. This lack in generalizability, however, does not impact our results regarding the differences in variance

explained by different scoring methods relative to each other, which was our main objective in conducting these analyses.

4.2 Conclusions

In summary, this study found that indices of alcohol consumption and AUD symptoms can be validly scaled along the same underlying continuum of alcohol involvement. Notably, alcohol involvement indices augmented information obtained by AUD symptoms alone in two ways: by measuring gaps between regions measured by AUD symptoms at the high severity end, and by providing information on the lower end of alcohol severity, which is not well-scaled by AUD symptoms. As such, our findings indicate that using alcohol consumption indices in addition to AUD symptoms increases measurement precision, which we found to have a noticeably positive effect on predicting later alcohol involvement. Questions remain regarding the sensitivity of consumption indices to differential item functioning. Further research is needed that examines the differential item functioning of consumption items of both low and high severity in different demographic groups, and particular in respect to age.

Acknowledgments

This study was supported by grants R21 AA016524 and T32 AA007459 from the National Institute on Alcohol Abuse and Alcoholism.

References

- Achenbach TM. Some needed changes in DSM-V: But what about children? *Clinical Psychology: Science and Practice*. 2009; 16:50–53.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association; Washington, DC: 1981.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association; Washington, DC: 1994.
- Birnbaum, A. Some latent trait models. In: Lord, FM.; Novick, MR., editors. *Statistical Theory of Mental Test Scores*. Addison-Wesley; Reading, MA: 1968. p. 397–472.
- Borges G, Ye Y, Bond J, Cherpitel CJ, Cremonte M, Moskalewicz J, Swiatkiewicz G, Rubio-Stipec M. The dimensionality of alcohol use disorders and alcohol consumption in a cross-national perspective. *Addiction*. 2010; 105:240–254. [PubMed: 20078482]
- Bucholz, KK. In: Bucholz, KK.; Hasin, DS., editors. Should craving be added as a criterion for alcohol use disorders in DSM-V?; *Alcohol Use Disorders in DSM-V: A Progress Report, Symposium presented at the 32nd Annual Scientific Meeting of the Research Society on Alcoholism*; June 20–24; San Diego, CA. 2009.
- Carney MA, Tennen H, Affleck G, Del Boca FK, Kranzler HR. Levels and patterns of alcohol consumption using Timeline Follow-Back, Daily Diaries and Real-Time “Electronic Interviews”. *J Stud Alcohol*. 1998; 59:447–454. [PubMed: 9647427]
- Chung T, Martin CS. Classification and course of alcohol problems among adolescents in addictions treatment programs. *Alcohol Clin Exp Res*. 2001; 25:1734–1742. [PubMed: 11781506]
- Chung T, Martin CS. What were they thinking? Adolescents’ interpretations of DSM-IV alcohol dependence symptom queries and implications for diagnostic validity. *Drug Alcohol Depend*. 2005; 80:191–200. [PubMed: 15894432]
- Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT’s unidimensionality assumption. *Qual Life Res*. 2009; 18:447–460. [PubMed: 19294529]
- Dawson DA. Methodological issues in measuring alcohol use. *Alcohol Res Health*. 2003; 27:18–29. [PubMed: 15301397]

- Dawson DA, Saha TD, Grant BF. A multidimensional assessment of the validity and utility of alcohol use disorder severity as determined by item response theory models. *Drug Alcohol Depend.* 2010; 107:31–38. [PubMed: 19782481]
- Draba, RE. The identification and interpretation of item bias. Educational Statistics Laboratory, University of Chicago; 1977.
- Edwards G, Gross MM. Alcohol dependence, provisional description of a clinical syndrome. *Br Med J.* 1976; 1:1058–1061. [PubMed: 773501]
- Faden VB, Fay MP. Trends in Drinking Among Americans Age 18 and Younger: 1975–2002. *Alcohol Clin Exp Res.* 2004; 28:1388–1395. [PubMed: 15365310]
- Gelhorn H, Hartman C, Sakai J, Stallings M, Young S, Rhee SH, Corley R, Hewitt J, Hopfer C, Crowley T. Toward DSM-V: An item response theory analysis of the diagnostic process for DSM-IV alcohol abuse and dependence in adolescents. *J Am Acad Child Adolesc Psychiatry.* 2008; 47:1329–1339. [PubMed: 18827724]
- Greenfield TK, Kerr WC. Alcohol measurement methodology in epidemiology: Recent advances and opportunities. *Addiction.* 2008; 103:1082–1099. [PubMed: 18422826]
- Guze SB, Tuason VB, Stewart MA, Picken B. The drinking history: A comparison of reports by subjects and their relatives. *Q J Stud Alcohol.* 1963; 24:249–260. [PubMed: 13951521]
- Harford TC, Yi H-y, Faden VB, Chen CM. The Dimensionality of DSM-IV Alcohol Use Disorders Among Adolescent and Adult Drinkers and Symptom Patterns by Age, Gender, and Race/Ethnicity. *Alcohol Clin Exp Res.* 2009; 33:868–878. [PubMed: 19320629]
- Hasin DS, Beseler CL. Dimensionality of lifetime alcohol abuse, dependence and binge drinking. *Drug Alcohol Depend.* 2009; 101:53–61. [PubMed: 19095379]
- Hoepfner BB, Stout RL, Jackson KM, Barnett NP. How good is fine-grained Timeline Follow-back data? Comparing 30-day TLFB and repeated 7-day TLFB alcohol consumption reports on the person and daily level. *Addict Behav.* in press.
- Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika.* 1965; 30:179–185. [PubMed: 14306381]
- Johnston, LD.; O'Malley, PM.; Bachman, JG.; Schulenberg, JE. Secondary school students (NIH Publication No. 09-7402). Vol. I. National Institute on Drug Abuse; Bethesda, MD: 2009. Monitoring the Future national survey results on drug use, 1975–2008.
- Kahler CW, Hoepfner BB, Jackson KM. A Rasch model analysis of alcohol consumption and problems across adolescence and young adulthood. *Alcohol Clin Exp Res.* 2009; 33:1–11. [PubMed: 18828798]
- Kahler CW, Strong DR. A Rasch model analysis of DSM-IV alcohol abuse and dependence items in the National Epidemiological Survey on Alcohol and Related Conditions. *Alcohol Clin Exp Res.* 2006; 30:1165–1175. [PubMed: 16792564]
- Kahler CW, Strong DR, Papandonatos GD, Colby SM, Clark MA, Boergers J, Niaura R, Abrams DB, Buka SL. Cigarette smoking and the lifetime alcohol involvement continuum. *Drug Alcohol Depend.* 2008; 93:111–120. [PubMed: 17964082]
- Kahler CW, Strong DR, Stuart GL, Moore TM, Ramsey SE. Item functioning of the alcohol dependence scale in a high-risk sample. *Drug Alcohol Depend.* 2003; 72:183–192. [PubMed: 14636973]
- Keyes KM, Geier T, Brant BF, Hasin DS. Influence of a drinking quantity and frequency measure on the prevalence and demographic correlates of DSM-IV alcohol dependence. *Alcohol Clin Exp Res.* 2009; 33:761–771. [PubMed: 19298332]
- Li TK, Hewitt BG, Grant BF. Is there a future for quantifying drinking in the diagnosis, treatment, and prevention of alcohol use disorders? *Alcohol Alcohol.* 2007; 42:57–63. [PubMed: 17307789]
- Linacre JM. Detecting multidimensionality: which residual data-type works best? *J Outcome Meas.* 1998; 2:266–283. [PubMed: 9711024]
- Linacre, JM. WINSTEPS Rasch Measurement: User's Manual. 2006. www.winsteps.com. accessed on
- Linacre JM, Wright BD. Reasonable mean-square fit values. *Rasch Measurement Transactions.* 1994; 8:370.
- Lord, FM. Applications of Item Response Theory to Practical Testing Problems. Erlbaum; Hillsdale, NJ: 1980.

- Maisto S, Conigliaro JC, Gordon AJ, McGinnis KA, Justice AC. An experimental study of the agreement of self-administration and telephone administration of the Timeline Followback interview. *J Stud Alcohol Drugs*. 2008; 69:468–471. [PubMed: 18432391]
- Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959; 22:719–748. [PubMed: 13655060]
- Martin CS, Chung T, Kirisci L, Langenbucher JW. Item response theory analysis of diagnostic criteria for alcohol and cannabis use disorders in adolescents: Implications for DSM-V. *J Abnorm Psychol*. 2006; 115:807–814. [PubMed: 17100538]
- National Institute on Alcohol Abuse and Alcoholism. Rethinking Drinking (NIH Publication #09-3770). National Institute on Alcohol Abuse and Alcoholism; Bethesda, MD: 2009.
- O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments & Computers*. 2000; 32:396–402.
- Perrine MW, Mundt JC, Searles JS, Lester LS. Validation of daily self-reported alcohol consumption using interactive voice response (IVR) technology. *J Stud Alcohol*. 1995; 56:487–490. [PubMed: 7475027]
- Proudfoot H, Baillie AJ, Teesson M. The structure of alcohol dependence in the community. *Drug Alcohol Depend*. 2006;81.
- Rasch, G. Probabilistic Models for Some Intelligence and Attainment Tests. Danish Institute for Educational Research; Copenhagen, Sweden: 1960.
- Saha TD, Chou SP, Grant BF. Toward an alcohol use disorder continuum using item response theory: Results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Psychol Med*. 2006; 36:931–941. [PubMed: 16563205]
- Saha TD, Stinson FS, Grant BF. The role of alcohol consumption in future classifications of alcohol use disorders. *Drug Alcohol Depend*. 2007; 89:82–92. [PubMed: 17240085]
- Salzberger T. The illusion of measurement: Rasch versus 2-PL. *Rasch Measurement Transactions*. 2002; 16:882.
- Smith, RM.; Miao, CY. Assessing unidimensionality for Rasch measurement. In: Wilson, M., editor. *Objective Measurement: Theory into Practice*. Ablex; Greenwich: 1994.
- Toll BA, Cooney NL, McKee SA, O'Malley SS. Correspondence between interactive voice response (IVR) and Timeline Followback (TLFB) reports of drinking behavior. *Addict Behav*. 2006; 31:726–731. [PubMed: 15975732]
- Tristán A. An Adjustment for Sample Size in DIF Analysis. *Rasch Measurement Transactions*. 2006; 20:1070–1071.
- Udry, JR. The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002 [machine-readable data file and documentation]. Carolina Population Center, University of North Carolina at Chapel Hill; Chapel Hill, NC: 2003.
- United States Department of Health and Human Services. Drug Abuse Treatment Outcome Study--Adolescent (DATOS-A). Inter-university Consortium for Political and Social Research; Ann Arbor, MI: 1993–1995. National Institute on Drug Abuse.
- World Health Organization. Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death. World Health Organization; Geneva, Switzerland: 1992.
- Wright BD. 3-PL or Rasch? *Rasch Measurement Transactions*. 1995; 9:408–409.
- Wright, BD.; Masters, GN. *Rating Scale Analysis*. MESA Press; Chicago: 1982.
- Zwick WR, Velicer WF. Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*. 1986; 99:432–442.

Figure 1.

Figure 1 illustrates the impact of DIF on the total test score level. For each subgroup, severity estimates of each score (i.e., 0–19) are plotted and compared within sex (A), race (B) and age (C). For all three comparisons, discrepancies in the severity of total test scores are most noticeable for low alcohol involvement scores. Alignment is best around a score of 4, which correspond roughly to a severity of -1.5 , ranging from -1.49 for Whites to -1.70 for African Americans. Slight departures from alignment are noticeable throughout medium to high scores, most noticeably between race groups.

Figure 2.

Test information functions are graphed for two sets of items: using AUD symptoms only, and using both AUD symptoms and consumption items. The severity of each item is denoted by a circle symbol and corresponding drop line, where black drop lines are used for consumption items and grey drop lines for AUD symptoms. To highlight the regions of the alcohol involvement continuum that consumption items contribute to, AUD symptoms are denoted on the “AUD symptoms only” test function, though they are part of both graphed functions. Two consumption items (i.e., “drunk” with $\theta = -9.10$ and “5+ drinks” with $\theta = -5.36$) are outside the graphed range.

Table 1
Items, percentages of lifetime endorsement, item response model parameters, and differential item functioning

Alcohol Involvement Item (<i>least to most severe</i>)	Type	%	Severity		Infit (MSQ)	PTBS	Differential Item Functioning (DIF)				
			EST	SE			Sex	WvB	WvH	BvH	Age
Drunk	cons.	80.0	-9.10	0.14	1.13	0.77					
Most drinks in a row: 5+	cons.	69.3	-5.36	0.09	0.64	0.89	** (f)	** (B)	** (H)		
Most drinks in a row: 12+	cons.	47.0	-1.59	0.05	1.29	0.70	** (f)	** (B)		* (B)	
5+ drinks per day once a week for 2+ months	cons.	35.5	-0.49	0.05	0.94	0.65		* (W)	* (W)		** (y)
14+/7+ drinks per week male/female	cons.	33.0	-0.30	0.05	0.96	0.64	* (m)	** (W)	** (W)	* (H)	
Larger/longer	dep.	32.6	0.01	0.05	0.86	0.62			** (H)		* (o)
Social/interpersonal Problems	abuse	28.9	0.39	0.05	1.15	0.54					
Hazardous use	abuse	25.8	0.50	0.05	0.93	0.56			* (B)		
Difficulty stopping before becoming intoxicated	dep.	24.6	0.60	0.05	0.93	0.55					
Tolerance	dep.	23.2	0.69	0.05	1.06	0.52					
Drunk for several days in a row	cons.	22.8	1.14	0.06	0.91	0.49					
Physical/psychological problems	dep.	18.5	1.26	0.06	0.83	0.50					
Needed so badly could not think of anything else	dep.	16.9	1.37	0.06	0.93	0.47	* (m)				
Time spent	dep.	16.2	1.47	0.06	0.85	0.47					
Activities given up	dep.	15.4	1.49	0.06	1.03	0.45					
5+ drinks every day for 2+ weeks	cons.	15.3	1.62	0.06	0.91	0.45					** (y)
Quit/control	dep.	14.2	2.00	0.06	0.93	0.40					** (o)
Made sure you had alcohol available over 1+ months	dep.	11.3	2.02	0.06	1.46	0.32					
Withdrawal	dep.	9.8	2.26	0.07	0.97	0.36					** (o)

Note :

* statistically significant and DIF > 0.50;

** statistically significant and DIF > 0.64; "m", "f", "W", "B", "H", "y", and "o" denote the group in the pairwise comparison for whom the item is more severe, where f=female ($n=885$, 26.2%), m=male ($n=2497$, 73.8%), W=White ($n=1745$, 51.6%), B=Black ($n=807$, 23.9%), H=Hispanic ($n=694$, 20.5%), y=younger (12–14 years, $n=629$, 18.6%), and o=older (15–18 years, $n=2753$, 81.4%)

Table 2

Subgroup specific item severity, DIF test statistics and effect size for items with DIF

Alcohol Involvement Item (<i>least to most severe</i>)	Subgroup Comparison	Severity		<i>t</i>	Effect Size
(C) Most drinks in a row: 5+					
Sex		−5.18 (m)	−4.49 (f)	−3.3	**
Race: White vs. Black		−5.57 (W)	−4.32 (B)	−5.5	**
Race: White vs. Hispanic		−5.57 (W)	−4.78 (H)	−3.2	**
(C) Most drinks in a row: 12+					
Sex		−1.86 (m)	−0.63 (f)	−9.9	**
Race: White vs. Black		−1.73 (W)	−0.97 (W)	−4.8	**
Race: Black vs. Hispanic		−0.97 (B)	−1.47 (B)	2.7	*
(C) 5+ drinks per day once a week for 2+months					
Race: White vs. Black		−0.23 (W)	−0.84 (B)	4.0	*
Race: White vs. Hispanic		−0.23 (W)	−0.76 (H)	3.7	*
Age		0.04 (y)	−0.53 (o)	3.9	*
(C) 14+/7+ drinks per week male/female					
Sex		−0.06 (m)	−0.70 (f)	5.2	**
Race: White vs. Black		0.12 (W)	−1.23 (B)	8.6	**
Race: White vs. Hispanic		0.12 (W)	−0.69 (H)	5.7	**
Race: Black vs. Hispanic		1.23 (B)	−0.69 (B)	−2.9	**
(D) Larger/longer					
Race: White vs. Hispanic		−0.42 (W)	0.15 (H)	−3.9	*
Age		−0.82 (y)	−0.12 (o)	−4.7	**
(A) Hazardous Use					
Race: White vs. Black		−0.37 (W)	0.36 (B)	−2.1	**
(D) Needed so badly could not think of anything else					
Sex		1.53 (m)	0.99 (f)	4.1	*
(C) 5+ drinks every day for 2+ weeks					
Age		2.31 (y)	1.49 (o)	4.3	**
(D) Quit/control					
Age		1.27 (y)	1.82 (o)	−3.3	*
(D) Withdrawal					
Age		1.73 (y)	2.53 (o)	−4.4	**

Note :

* statistically significant and DIF > 0.50;

** statistically significant and DIF > 0.64;

C=consumption, D=dependence, A=abuse, f=female(n=885, 26.2%), m=male (n=2497,73.8%), W=White (n=1745, 51.6%), B=Black (n=807, 23.9%), H=Hispanic (n=694, 20.5%), y=younger (12–14 years, n=629, 18.6%), and o=older (15–18 years, n=2753, 81.4%)

Table 3

Predictive capability of different scoring methods for 12-month alcohol involvement

12-Month Outcome Predictor	EST	SE	Wald χ^2	p	Pseudo R ²	Model
Any alcohol in past year						Logistic
Sum Score: AUD and consumption items	-0.16	0.06	6.3	0.012	0.090	
Sum Score: AUD symptoms	-0.11	0.06	3.8	0.051	0.088	
Diagnosis: DSM-IV Dependence	-0.06	0.05	1.4	0.234	0.083	
Number of days drunk in past year						Poisson
Sum Score: AUD and consumption items	0.35	0.01	3192.3	<0.001	0.181	
Sum Score: AUD symptoms	0.28	0.01	2595.8	<0.001	0.150	
Diagnosis: DSM-IV Dependence	0.21	0.01	1768.7	<0.001	0.063	
Ever drunk for several days in a row in past year						Logistic
Sum Score: AUD and consumption items	-0.50	0.09	30.3	<0.001	0.083	
Sum Score: AUD symptoms	-0.39	0.08	23.5	<0.001	0.070	
Diagnosis: DSM-IV Dependence	-0.36	0.07	26.2	<0.001	0.070	

Note: All predictors were standardized prior to analysis. All analyses include sex, race, age and “having been in a controlled environment at some point during the past 12-month” as covariates. Pseudo R² values represent Nagelkerke’s R² for logistic models, and deviance-based R² (Cameron, 1996) for the Poisson model.